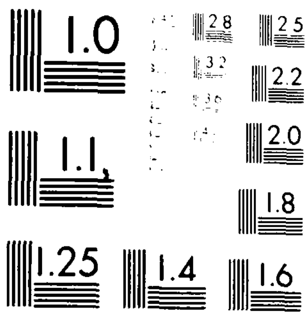
AD-A131 784 BAYESIAN STATISTICAL ESTIMATION IN THE FINITE STATE MARKOV RENEWAL PROCESS(U) CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS G T DUNCAN DEC 77 1/1

UNCLASSIFIED TR-139 N00014-77-C-0095 F/G 12/1 NL

END
DATE
FILMED
9 83
DTIC



MICROCOPY RESOLUTION TEST CHART

AD A131784

DEPARTMENT

OF

STATISTICS

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

BAYESIAN STATISTICAL ESTIMATION IN THE FINITE STATE MARKOV RENEWAL PROCESS

George T. Duncan
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA. 15213

Technical Report No. 139

December, 1977

An earlier version of this report was presented at the National Meeting of the Operations Research Society of America, San Francisco, California, May 9, 1977. The author wishes to thank Mr. Che Fang for his assistance with this work and Professor Morris H. DeGroot for his comments.

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by ONR under Contract No. N00014-77-C-0095.

Abstract

Bayesian estimators for the parameters of the finite state Markov renewal process are developed, both for the waiting time distributions and the transition probabilities.

KEY WORDS

Markov, Renewal process, Bayesian, Estimation, Waiting times, Transition probabilities

1. INTRODUCTION

This paper develops Bayesian statistical estimation procedures for the finite state Markov renewal process. The general case is treated where uncertainty exists about both the waiting time distributions and the transition probabilities. This work extends the Bayesian results of Martin (1967), who only considers the Markov chain case, and Brock (1971), who assumes that the waiting time distributions are known. Moore and Pyke (1968) deal only with classical estimation methods. The sampling schemes considered are either (1) to observe n transitions and their associated waiting times, or, more generally, ~~(2)~~ to observe the process for some time T, where T is not necessarily a transition time.

2. THE MARKOV RENEWAL PROCESS MODEL

The Markov renewal process is a convenient and workable generalization of both a Markov process and a renewal process, incorporating the essential features of each. It is closely related to the semi-Markov process independently investigated by Lévy [1954], Smith [1954] and Takács [1954]. Essentially, a finite state semi-Markov process is a stochastic process modelling moves among a finite number of states with the successive states visited forming a Markov chain and the length of stay in a given state being a random variable, the distribution function of which may depend on this (origin) state as well as on the one to be visited next. The finite state semi-Markov process can then be thought of as a Markov chain for which the time scale has been randomly transformed.

As noted by Pyke [1961a] the semi-Markov process is equivalent to the Markov renewal process which records at each time t the number of times an entity has visited each of the possible states up to time t, if the entity moves from state to state according to a Markov chain, and if the time required for each successive move is a random variable whose distribution function may depend on the two states between which the move is being made. An early application of the semi-Markov process was pursued by Cane [1959]. Its application to the social science phenomena of social mobility and migration is outlined by Ginsberg [1971, 1972a,b] and it is reviewed briefly by Bartholomew [1973]. Statistical inference questions about the transition probabilities are considered by Moore and Pyke [1968]. Building on the work of Martin [1967], Brock [1971] examines a Bayesian procedure for inference about the transition probabilities.

We shall deal with a finite state space, labelled $S = \{1,2,\ldots,s\}$, and we let $Q_{ij}(t)$ denote the probability that after making a transition into state i, the process next makes a transition into state j, in an amount of time less than or equal to t. Note explicitly that the Markov renewal model allows for the possibility of transitions from state i to state i. Such transitions may or may not have substantive meaning in any particular application. For example, in migration studies a move within a particular geographic region (state i) might signal a transition. On the other hand, in reliability studies the states usually reflect the operating status of a system, and a "within state" transition may have no physical meaning. In this later case $Q_{ii}(t) = 0$. But, in general, we must have $Q_{ij}(t) \geq 0$, $i,j = 1,\ldots,s$; $t \geq 0$. Let $P_{ij} = Q_{ij}(\infty) = \lim_{t\to\infty} Q_{ij}(t)$ and note that $\sum_{j=1}^{s} P_{ij} = 1$, $i=1,\ldots,s$. If $P_{ij} > 0$, let

$$F_{ij}(t) = \frac{Q_{ij}(t)}{P_{ij}} \tag{1}$$

(If $P_{ij} = 0$, let $F_{ij}(t)$ be arbitrary.)

With this notation P_{ij} represents the probability that the next transition will be into state j, given that the process has just entered state i, and $F_{ij}(t)$ represents the conditional probability that a transition will take place within an amount of time t, given that the process has just entered i and will next enter j. When i is entered, the next state is chosen according to the transition probabilities P_{ij}; then given that the state chosen is j, the time until transition has a distribution $F_{ij}(\cdot)$. These quantities P_{ij} and $F_{ij}(t)$ are estimable from data through the transition frequencies and observed waiting times between the various transitions. We begin the estimation process in the next section.

3. THE LIKELIHOOD FUNCTION

If the Markov renewal process is observed through time T, during which n transitions take place, the data will have the form

$$(X_0, T_1, X_1, \ldots, T_n, X_n, W) = (x_0, t_1, x_1, \ldots, t_n, x_n, w)$$

where X_o is the state initially occupied, X_i is the state occupied after the i^{th} transition, and T_i is the waiting time between the $(i-1)^{st}$ and i^{th} transition. The likelihood function is then proportional to

$$L = \left[\prod_{i=1}^{n} q_{x_{i-1},x_i}(t_i)\right]\left[1 - \sum_{j=1}^{s} Q_{x_n j}(w)\right],$$

where $q_{ij}(t)$ is the probability density function corresponding to $Q_{ij}(t)$ and

$$w = T - \sum_{i=1}^{n} t_i .$$

Using equation (1), L can be rewritten as

$$L = \left[\prod_{i=1}^{n} f_{x_{i-1},x_i}(t_i)\right]\left[\prod_{i=1}^{n} P_{x_{i-1},x_i}\right]\left[1 - \sum_{j=1}^{s} F_{x_n j}(w)P_{x_n j}\right],$$

where $f_{ij}(t)$ is the probability density function corresponding to $F_{ij}(t)$. This is the same as the likelihood function derived by Moore and Pyke (1968).

In most practical applications there will not be enough data to adequately estimate the waiting time distribution functions F directly, with no restrictions placed on F. Thus it will not usually be feasible to estimate F through the empirical distribution function. Instead it will be necessary to restrict the class of distribution functions to some general parametric family and then estimate the indexing parameters of this family. It will also be true that in most practical applications F will be absolutely continuous. We therefore assume that F is a member of some parametric family of continuous distributions, indexed by the parameter vector $\underset{\sim}{\theta}$. We will then write as $F_{ij}^{\underset{\sim}{\theta}}$ the distribution function of the waiting times between transitions from state i to state j, with $f_{ij}^{\underset{\sim}{\theta}}$ denoting the corresponding density function. Then letting n_{ru} denote the number of observed transitions from state r to state u, L can be further rewritten as

$$L = L_{\underset{\sim}{\theta}} L_{\underset{\sim}{p}} - \sum_{j=1}^{s} \left[L_{\underset{\sim}{\theta}} F_{x_n j}^{\underset{\sim}{\theta}}(w) \right] L_{\underset{\sim}{p}} P_{x_n j} , \tag{2}$$

where

$$L_{\underset{\sim}{\theta}} = \prod_{r,u=1}^{s} \prod_{k=1}^{n_{ru}} f_{ru}^{\underset{\sim}{\theta}}(t_{ru(k)}) ,$$

with $t_{ru(k)}$ being the observed waiting time between the $(k-1)^{st}$ and k^{th} transition from state r to state u, and

$$L_{\underset{\sim}{p}} = \prod_{r,u=1}^{s} P_{ru}^{n_{ru}} .$$

Thus the sufficient statistic for this model is a vector $\underset{\sim}{Z} = (W, \underset{\sim}{Z}^{(1)}, \underset{\sim}{Z}^{(2)})$, where $Z^{(1)}$ denotes the vector of observed waiting times, $t_{ru}(k)$, between pairs of states, and $Z^{(2)}$ denotes the vector of transition counts n_{ru}.

4. THE FORM OF BAYES ESTIMATORS

We consider a Bayesian treatment of the estimation problem. This is desirable because of the flexibility of the Bayesian approach in incorporating varying amounts of prior information and its success in handling the problem of limited amounts of relevant sampling data.

We now derive Bayes estimators of $\underset{\sim}{P}$ and $\underset{\sim}{\theta}$ using squared error loss. Thus we require the mean of the posterior distribution of $\underset{\sim}{P}$ and $\underset{\sim}{\theta}$. Let the prior joint density of $\underset{\sim}{P}$ and $\underset{\sim}{\theta}$ be denoted by π. Then the Bayes estimator of $\underset{\sim}{P}$ is

$$E(\underset{\sim}{P}|\underset{\sim}{z}) = \frac{\int \underset{\sim}{P} L \pi d\underset{\sim}{P} d\underset{\sim}{\theta}}{\int L \pi d\underset{\sim}{P} d\underset{\sim}{\theta}} . \tag{3}$$

The form of the Bayes estimator will simplify substantially if we assume prior independence of $\underset{\sim}{P}$ and $\underset{\sim}{\theta}$, i.e., that π can be written in the form

$$\pi = \pi(\underset{\sim}{P}) \cdot \pi(\underset{\sim}{\theta}).$$

By using equation (2) we can rewrite the normalizing constant, $\int L \pi d\underset{\sim}{P} d\underset{\sim}{\theta}$, as follows:

First, to simplify notation, let

$$F^{\theta}_{x_n,j}(w) = F^{\theta}_{j} \quad \text{and} \quad P_{x_n,j} = P_j .$$

Then

$$\int L \pi d\underset{\sim}{P} d\underset{\sim}{\theta} =$$

$$\int L_{\underset{\sim}{\theta}} L_{\underset{\sim}{P}} \pi(\underset{\sim}{P}) \pi(\underset{\sim}{\theta}) d\underset{\sim}{P} d\underset{\sim}{\theta} - \sum_{j=1}^{s} \int L_{\underset{\sim}{\theta}} \widetilde{F^{\theta}_{j}} L_{\underset{\sim}{P}} P_j \pi(\underset{\sim}{P}) \pi(\underset{\sim}{\theta}) d\underset{\sim}{P} d\underset{\sim}{\theta}$$

=

$$\left[\int L_{\underset{\sim}{\theta}}\,\pi(\underset{\sim}{\theta})d\underset{\sim}{\theta}\right]\left[\int L_{\underset{\sim}{P}}\,\pi(\underset{\sim}{P})d\underset{\sim}{P}\right]-\sum_{j=1}^{s}\left[\int L_{\underset{\sim}{\theta}}F_j^{\underset{\sim}{\theta}}\,\pi(\underset{\sim}{\theta})d\underset{\sim}{\theta}\right]\left[\int L_{\underset{\sim}{P}}\,P_j\,\pi(\underset{\sim}{P})d\underset{\sim}{P}\right]. \qquad (4)$$

Similarly,

$$\int \underset{\sim}{P}\,L\,\pi d\underset{\sim}{P}\,d\underset{\sim}{\theta} =$$

$$\left[\int L_{\underset{\sim}{\theta}}\,\pi(\underset{\sim}{\theta})d\underset{\sim}{\theta}\right]\left[\int \underset{\sim}{P}L_{\underset{\sim}{P}}\pi(\underset{\sim}{P})d\underset{\sim}{P}\right]-\sum_{j=1}^{s}\left[\int L_{\underset{\sim}{\theta}}F_j^{\underset{\sim}{\theta}}\pi(\underset{\sim}{\theta})d\underset{\sim}{\theta}\right]\left[\int \underset{\sim}{P}L_{\underset{\sim}{P}}\,P_j\pi(\underset{\sim}{P})d\underset{\sim}{P}\right]. \qquad (5)$$

To obtain the Bayes estimator of $\underset{\sim}{\theta}$ we need to compute the posterior expectation,

$$E\underset{\sim}{\theta} = \frac{\int \underset{\sim}{\theta} L\pi d\underset{\sim}{P}d\underset{\sim}{\theta}}{\int L\pi d\underset{\sim}{P}d\underset{\sim}{\theta}} \quad . \qquad (6)$$

The numerator of equation (6) can be rewritten as

$$\int \underset{\sim}{\theta} L\pi d\underset{\sim}{P}d\underset{\sim}{\theta} =$$

$$\left[\int \underset{\sim}{\theta} L_{\underset{\sim}{\theta}}\pi(\underset{\sim}{\theta})d\underset{\sim}{\theta}\right]\left[\int L_{\underset{\sim}{P}}\pi(\underset{\sim}{P})d\underset{\sim}{P}\right] - \sum_{j=1}^{s}\left[\int \underset{\sim}{\theta} L_{\underset{\sim}{\theta}}F_j^{\underset{\sim}{\theta}}\pi(\underset{\sim}{\theta})d\underset{\sim}{\theta}\right]\left[\int L_{\underset{\sim}{P}}P_j\pi(\underset{\sim}{P})d\underset{\sim}{P}\right]. \qquad (7)$$

Note that under the assumption of independent prior distributions for $\underset{\sim}{P}$ and $\underset{\sim}{\theta}$ the calculation of Bayes estimates involves separate evaluations of integrals with respect to $\underset{\sim}{\theta}$ and integrals with respect to $\underset{\sim}{P}$. Also note the substantial simplication that occurs when the terminal time T happens to be a transition time, i.e., when $w=0$. In this case the Bayes estimators are just

$$\underset{\sim}{P}^* = E(\underset{\sim}{P}|\underset{\sim}{Z}^{(2)}) = \frac{\int \underset{\sim}{P}L_{\underset{\sim}{P}}\pi(\underset{\sim}{P})d\underset{\sim}{P}}{\int L_{\underset{\sim}{P}}\,\pi(\underset{\sim}{P})d\underset{\sim}{P}},$$

and

$$\underset{\sim}{\theta}^* = E(\underset{\sim}{\theta}|\underset{\sim}{Z}^{(1)}) = \frac{\int \underset{\sim}{\theta} L_{\underset{\sim}{\theta}} \pi(\underset{\sim}{\theta}) d\underset{\sim}{\theta}}{\int L_{\underset{\sim}{\theta}} \pi(\underset{\sim}{\theta}) d\underset{\sim}{\theta}} .$$

In the general case where $w \geq 0$ the Bayes estimators take a form adjusted away from $\underset{\sim}{P}^*$ and $\underset{\sim}{\theta}^*$. Specifically, using equations (4) and (5), we can write the Bayes estimator $\hat{\underset{\sim}{P}}$ as

$$\hat{\underset{\sim}{P}} = \frac{\underset{\sim}{P}^* - \sum_{j=1}^{s} \alpha_j E(\underset{\sim}{P} P_j|\underset{\sim}{Z}^{(2)})}{1 - \sum_{j=1}^{s} \alpha_j \beta_j} , \tag{8}$$

where

$$\alpha_j = E(F_j^{\underset{\sim}{\theta}}|\underset{\sim}{Z}^{(1)})$$

and

$$\beta_j = E(P_j|\underset{\sim}{Z}^{(2)}).$$

The Bayes estimator $\hat{\underset{\sim}{\theta}}$ can be written using equations (4) and (7) as

$$\hat{\underset{\sim}{\theta}} = \frac{\underset{\sim}{\theta}^* - \sum_{j=1}^{s} \beta_j E(\underset{\sim}{\theta} F_j^{\underset{\sim}{\theta}}|\underset{\sim}{Z}^{(1)})}{1 - \sum_{j=1}^{s} \alpha_j \beta_j} . \tag{9}$$

The practical use of the Bayes estimators $\hat{\underset{\sim}{P}}$ and $\hat{\underset{\sim}{\theta}}$ is limited by the fact that integration of functions of $\underset{\sim}{\theta}$ involving $F^{\underset{\sim}{\theta}}$ is required, and $F^{\underset{\sim}{\theta}}$ is generally a quite complicated function of $\underset{\sim}{\theta}$. An alternative pair of estimators

$\tilde{P}$ and $\tilde{\theta}$ can be derived by replacing F^{θ} in equations (8) and (9) with the constant (over θ), F^{θ^*}. Writing then $F_j^{\theta^*}$ as just α_j^*, we have the simplified estimators,

$$\tilde{\theta} = E(\theta | Z^{(1)}) \qquad (= \theta^*)$$

and

$$\tilde{P} = \frac{P^* - \sum_{j=1}^{s} \alpha_j^* E(\beta P_j | Z^{(2)})}{1 - \sum_{j=1}^{s} \alpha_j^* \beta_j} . \tag{10}$$

5. MODELS FOR WAITING TIME DISTRIBUTIONS

We seek parametric families of continuous positive random variables which are sufficiently rich to represent a wide range of possible waiting time distributions, while at the same time being analytically tractable in terms of the Bayesian estimation procedures developed in Section 4. Two families of distributions, each specified by 2 parameters, immediately suggest themselves for this purpose: the lognormal and the gamma. They are discussed in Johnson and Kotz [1970] and Mann, Schafer, and Singpurwalla [1974]. A random variable T is said to have a (two-parameter) lognormal distribution if the natural logarithm of T has a normal distribution (see Aitchison and Brown [1957]). Thus the distribution of T is determined by the mean μ and variance of σ^2 of the normal distribution. When σ^2 is small the distribution of T itself will be not unlike that of a normal distribution. The gamma family specifies that the probability density function of T would be given by

$$f_T(t) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}$$

for $t>0$ where α and β are positive parameters. Here again it is possible to transform the random variable T to a very good approximation of normality for only moderately large α. The Wilson Hilferty [1931] approximation can be used to show that $T^{1/3}$ has approximately a normal distribution with mean $(\alpha/\beta)^{1/3}(1-1/9\alpha)$ and variance $(\alpha/\beta)^{2/3}/9\alpha$.

It also should be noted that the distribution of $\log T$ is more nearly normal than the distribution of T, so that a logarithmic transformation to normality is ideal in the lognormal distribution case and helps in the gamma distribution case, while a cube root transformation is best in the gamma distribution case.

From a strictly empirical viewpoint it may be desirable to use the data to choose a transformation to normality from among the power transformations considered by Box and Cox [1964]. These include the cube root transformation; the logarithmic transformation is a limiting case. Thus we see that it will generally be possible to transform the waiting time data to achieve approximate normality.

Therefore we assume that it is possible to deal with transformed values U of the original waiting times T, these transformed values U having approximate normal distributions. The problem of statistical inference about the distribution of waiting times then becomes one of inference about the means and variances of the normal random variables U.

5. BAYES ESTIMATES OF THE WAITING TIME DISTRIBUTIONS

Suppose that it is possible to transform the vector of all waiting times $\underset{\sim}{T}$ $(s^2 \times 1)$ to a multivariate normal random vector $\underset{\sim}{U}$ $(s^2 \times 1)$. We seek to estimate the mean vector $\underset{\sim}{\mu}$ $(s^2 \times 1)$ and the variance matrix $\underset{\sim}{\Sigma}$ $(s^2 \times s^2)$ of $\underset{\sim}{U}$.

Assume that a prior distribution for $\underset{\sim}{\mu}$ and the <u>precision</u> matrix $\underset{\sim}{R} = \underset{\sim}{\Sigma}^{-1}$ is chosen from the natural conjugate family of normal-Wishart distributions. We then assume that the prior distribution has the following structure: The conditional distribution of $\underset{\sim}{\mu}$ given that $\underset{\sim}{R} = \underset{\sim}{r}$ is a multivariate normal distribution with mean vector $\underset{\sim}{\mu}_0$ $(s^2 \times 1)$ and precision matrix $\nu_0 \underset{\sim}{r} (\nu_0 > 0)$, and the marginal distribution of $\underset{\sim}{R}$ is a Wishart distribution with $\alpha_0 (\alpha_0 > s^2 - 1)$ degrees of freedom and positive definite precision matrix $\underset{\sim}{\tau}$ $(s^2 \times s^2)$. Let $\underset{\sim}{U}_1, \ldots, \underset{\sim}{U}_n$ be a random sample distributed like $\underset{\sim}{U}$.

Under these circumstances the essential basis for inference is contained in the following theorem.

<u>Theorem</u> (See, e.g., DeGroot (1970;p.178).)

The posterior joint distribution of $\underset{\sim}{\mu}$ and $\underset{\sim}{R}$ when $\underset{\sim}{U}_i = u_i$ $(i=1,\ldots,n)$ is as follows: The conditional distribution of $\underset{\sim}{\mu}$ when $\underset{\sim}{R} = \underset{\sim}{r}$ is a multivariate normal distribution with mean vector $\underset{\sim}{\mu}_1$ and precision matrix $(\nu_1)\underset{\sim}{r}$, where $\nu_1 = \nu_0 + n$ and

$$\underset{\sim}{\mu}_1 = \frac{\nu_0 \underset{\sim}{\mu}_0 + n\overline{\underset{\sim}{U}}}{\nu_1} .$$

The marginal distribution of $\underset{\sim}{R}$ is a Wishart distribution with $\alpha_1 = \alpha_0 + n$ degrees of freedom and precision matrix τ_1, where

$$\underset{\sim}{T}_1 = \underset{\sim}{T}_o + \underset{\sim}{S} + \frac{\nu_o n}{\nu_1}(\underset{\sim}{\mu}_o - \bar{\underset{\sim}{U}})(\underset{\sim}{\mu}_o - \bar{\underset{\sim}{U}})',$$

and

$$\underset{\sim}{S} = \sum_{i=1}^{n}(\underset{\sim}{U}_i - \bar{\underset{\sim}{U}})(\underset{\sim}{U}_i - \bar{\underset{\sim}{U}})'.$$

If we now define a multivariate quadratic loss function of the form

$$L(\underset{\sim}{\mu}, \tilde{\underset{\sim}{\mu}}) = (\underset{\sim}{\mu} - \tilde{\underset{\sim}{\mu}})'(\underset{\sim}{\mu} - \tilde{\underset{\sim}{\mu}}),$$

we can obtain the Bayes estimator of $\underset{\sim}{\mu}$ as the mean of the posterior distribution of $\underset{\sim}{\mu}$, namely,

$$\tilde{\underset{\sim}{\mu}} = \underset{\sim}{\mu}_1 .$$

Also for a loss function of the form

$$L(\underset{\sim}{\Sigma}, \tilde{\underset{\sim}{\Sigma}}) = tr(\underset{\sim}{\Sigma} - \tilde{\underset{\sim}{\Sigma}})^2,$$

the Bayes estimator for $\underset{\sim}{\Sigma}$ is

$$\tilde{\underset{\sim}{\Sigma}} = \frac{\underset{\sim}{T}_1}{\alpha + n - 2(s^2 - 1)}$$

(See, e.g., Press (1972;pp. 168ff).).

7. BAYES ESTIMATES OF THE TRANSITION PROBABILITIES

We consider a Bayesian approach to estimation of the transition probabilities $\underset{\sim}{P}$ in the Markov renewal process. Martin [13] developed a methodology for Bayesian inference about the transition probabilities of a Markov chain. He noted that in this case the observed transition counts in each row of $\underset{\sim}{P}$ come from a multinomial distribution with cell probabilities depending in a particular way on $\underset{\sim}{P}$. A <u>matrix</u> <u>beta</u>

distribution is then postulated as appropriate for a prior distribution. This is just the distribution of s independent Dirichlet random vectors (see Johnson and Kotz (1972;p.233). The matrix beta distribution is, like the Dirichlet, closed under sampling so that the posterior distribution is also a member of the matrix beta family.

Specifically, the matrix beta distribution is determined by a matrix parameter $\underset{\sim}{M}$. If $\underset{\sim}{M}_0$ is taken to be the matrix parameter specifying the prior distribution and a matrix of frequency counts $\underset{\sim}{N}$ is observed, then the posterior distribution of $\underset{\sim}{P}$ is matrix beta with matrix parameter $\underset{\sim}{M}=\underset{\sim}{M}_0+\underset{\sim}{N}$. Under quadratic loss the Bayes estimator of $\underset{\sim}{P}$ is then given by the posterior expectation of $\underset{\sim}{P}$. But the expectation of a random matrix $\underset{\sim}{P}$ having a matrix beta distribution with matrix parameter $\underset{\sim}{M}=[m_{ij}]$ is simply,

$$E\underset{\sim}{P} = [\frac{m_{ij}}{m_{i+}}] \tag{11}$$

where $m_{i+} = \sum_{j=1}^{s} m_{ij}$ (See Martin [1967; p. 17].)

This same approach is followed by Brock [1971] for the Markov renewal process when the waiting time distribution are assumed known.

For Bayesian estimation in the general Markov renewal process, we require, according to equation (10), two specific conditional expectations, $E(\underset{\sim}{P}|\underset{\sim}{Z}^{(2)})$ and $E(\underset{\sim}{P}P_{x_n,j}|\underset{\sim}{Z}^{(2)})$.

The first conditional expectation is available directly from equation (11). The second conditional expectation can be computed as follows: We basically require, for $\underset{\sim}{P}$ having a matrix beta distribution with matrix parameter $M = [m_{ru}]$, the expectation

$$E(\underset{\sim}{P}P_{ij}) = [EP_{ru}P_{ij}].$$

But

$$EP_{ru}P_{ij} = \mathrm{Cov}(P_{ru},P_{ij}) + EP_{ru}\ EP_{ij}.$$

Now from Martin [1967;p.18] we have

$$\begin{aligned}
\mathrm{Cov}(P_{ru},P_{ij}) &= -\frac{m_{ru}m_{ij}}{(m_{r+})^2(m_{i+}+1)}, \quad i=r, j\neq u, \\
&= \frac{m_{ru}(m_{i+}-m_{ij})}{(m_{r+})^2(m_{i+}+1)}, \quad i=r, j=u, \\
&= \quad 0 \quad , \text{ otherwise.}
\end{aligned}$$

Therefore, making use of equation (11) and simplifying, we have

$$\begin{aligned}
E(P_{ru}P_{ij}) &= \frac{m_{ru}m_{ij}}{m_{r+}(m_{i+}+1)}, \quad i=r, j\neq u, \\
&= \frac{m_{ru}(m_{ij}+1)}{m_{r+}(m_{i+}+1)}, \quad i=r, j=u, \qquad (12) \\
&= \frac{m_{ru}m_{ij}}{m_{r+}m_{i+}}, \quad i\neq r.
\end{aligned}$$

Substituting the calculated values in equations (11) and (12) into the estimator given in equation (10) we obtain after simplification,

$$P_{ru} = \frac{m_{ru}}{m_{r+}} \quad , \qquad r \neq x_n$$
$$\qquad = a_u \frac{m_{ru}}{m_{r+}} \quad , \qquad r = x_n , \tag{13}$$

where

$$a_u = \frac{1 - \sum\limits_{j \neq u} \alpha_j^* \dfrac{m_{x_n j}}{m_{x_n +} + 1} - \alpha_u^* \dfrac{m_{x_n u} + 1}{m_{x_n +} + 1}}{1 - \sum\limits_{j} \alpha_j^* \dfrac{m_{x_n j}}{m_{x_n +}} .} \quad .$$

It is easy to verify directly that with the matrix beta prior distribution either the Bayes estimates $\hat{P}$ or their approximations $\tilde{P}$ have row sums equal to one and have non-negative entries. Hence either $\hat{P}$ or $\tilde{P}$ can be used as a legitimate transition probability matrix.

8. A SIMULATION EXAMPLE

This section illustrates the application of the statistical techniques developed in the previous section to simulated data. The use of simulated data is advantageous at this point because a direct comparison can be made between results based on estimated process parameters and true process parameters.

Data from a 3-state Markov renewal process with log-normally distributed interarrival times were simulated on an IBM 360/67 computer at Carnegie-Mellon University. The transition probability matrix was

$$\underset{\sim}{P} = \begin{bmatrix} .3 & .6 & .1 \\ .2 & .5 & .3 \\ .1 & .3 & .6 \end{bmatrix} .$$

To give concreteness to this example, imagine that state $i=1,2,3$ indicates the number of projects that an analyst has been assigned to work on simultaneously. Often when a project is completed, it will be replaced immediately with another one. Hence there is high probability of an interstate move. The analyst is never assigned more than three projects and always is assigned at least one. The length of time a project will take to complete is a random variable. It is possible to complete two projects simultaneously and two additional projects might be assigned to the analyst currently working on one. The 9 waiting time distributions between transitions are taken to be lognormally distributed. The means μ and precisions τ (reciprocal of the variances) of these distributions are given, correspondingly to the entries of $\underset{\sim}{P}$, as the first line of each of 9 cells in Table 1, below.

TABLE 1

True Waiting Time Distribution Parameters

(1, 1) 4.5 2.7	(1, 1/4) 20.1 2.7	(2, 1/9) 665.1 7.4
(1, 1/9) 244.7 2.7	(2,1) 12.2 7.4	(2, 1/4) 54.6 7.4
(2, 1/9) 665.1 7.4	(2, 1/9) 665.1 7.4	(1,1) 4.5 2.7

Thus, for example, the logarithm of the time between a transition from state 2 to state 3 was normally distributed with a mean of 2 and a standard deviation of 2. The expected interarrival times T can be calculated from

$$ET = \exp(\mu + \frac{1}{2\tau})$$

and are given as the second line of each cell in Table 1. Since the lognormal distribution is skew the median interarrival times exp (μ) differ from the mean interarrival times and are given as the third line of each cell in Table 1. In our illustration, the units of T might be days.

The Bayes procedures outlined in the previous sections require the specification of the parameters of the prior distributions. There are 9 waiting time distributions, each of which requires that 4 parameters, μ_0, ν_0, α_0, and $\beta_0 = \frac{1}{2}\tau_0$ of the normal-gamma conjugate prior distribution be set. We will illustrate the technique with independent and identical prior distributions on all waiting time distributions. There is, of course, no reason why this should necessarily be the case in a practical application; we do this only for ease of illustration. In order to give a reasonably diffuse prior specification, while not being too far off from the target, we choose our hyperparameters to be $\mu_0 = 1$, $\lambda_0 = 1$, $\alpha_0 = 1$, and $\beta_0 = 4$. This leads to prior (minimum mean squared error) estimates (no-data) of the means and precisions of $\hat{\mu} = \mu_0 = 1$ and $\hat{\tau}^{-1} = (\alpha_0/\beta_0)^{-1} = (1/4)^{-1}$. Prior estimated expected inter-arrival times can then be calculated as

$$\hat{ET} = \exp(\hat{\mu} + \frac{1}{2}\frac{1}{\hat{\tau}}) = 20.086$$

and the estimated median waiting time would be

$$e^{\hat{\mu}} = 2.718.$$

Note that these estimated means and medians are not Bayes with respect to squared error loss. They are only given to illustrate their approximate magnitude, since our real concern is with μ and τ^{-1}. Naturally since the actual expected waiting times vary substantially among themselves this estimate misses some of the expected waiting times by a substantial margin.

To complete the specification of the prior parameters we need to give the 3×3 matrix parameter $\underset{\sim}{M}_0$ of the matrix beta distribution for the transition probability matrix $\underset{\sim}{P}$. We choose

$$\underset{\sim}{M}_o = \begin{pmatrix} 5 & 3 & 2 \\ 2 & 6 & 2 \\ 2 & 3 & 5 \end{pmatrix}$$

which leads to prior estimates of $\underset{\sim}{P}$ as

$$\hat{\underset{\sim}{P}} = \begin{pmatrix} .5 & .3 & .2 \\ .2 & .6 & .2 \\ .2 & .3 & .5 \end{pmatrix} .$$

Note that it is not necessary for the row sums of $\underset{\sim}{M}_o$ to be 10, they could be any positive number.

We now wish to show how these prior estimates will be modified with data. Simulated data of 91 transitions (beginning in state 1), together with their associated lognormally distributed waiting times are used. The last state occupied was $x_n=2$ and the observed time in this state before termination was $w = 21.3$. The stopping rule used was to stop when the total observation time was 631.8. The "observed" transition counts, based on the simulation, are given by

$$\underset{\sim}{N} = \begin{pmatrix} 2 & 7 & 3 \\ 7 & 20 & 13 \\ 2 & 12 & 25 \end{pmatrix} .$$

Thus the maximum likelihood estimate of the transition probability matrix is

$$\begin{bmatrix} .167 & .583 & .250 \\ .175 & .500 & .325 \\ .051 & .308 & .641 \end{bmatrix} .$$

Table 2 gives the maximum likelihood estimates of the parameters of the lognormal distributions of waiting times. The format of the table corresponds exactly to that of Table 1. Thus the maximum likelihood estimate of the mean waiting time for a transition from state 2 to state 3 is 538.215, while a comparison with Table 1 shows the true mean waiting time to be 54.6.

TABLE 2

Maximum Likelihood Estimates of the Waiting Time Distribution Parameters

(1.401, 4.120) 4.583 4.059	(-0.139, 0.605) 1.989 .870	(3.346, 0.090) 12,107.2 46.805
(1.774, 0.538) 14.930 5.394	(2.013, 1.220) 11.278 7.486	(2.256, 0.124) 538.215 9.545
(-0.045, 6.331) 1.035 .956	(1.696, 0.208) 60.331 5.452	(1.078, 0.673) 6.178 2.939

Employment of the Bayesian methods discussed in Section 6 produced the estimates $\tilde{\underset{\sim}{\theta}}$ of the parameters of the waiting time distributions contained in Table 3. Table 3 has the same format at Tables 1 and 2. We have simply displayed the estimated expected waiting times as $\exp(\tilde{\mu} + \frac{1}{2}\,\tilde{\tau}^{-1})$ and the estimated median waiting times as $\exp(\tilde{\mu})$, since our interest centers on minimum mean squared error estimation of μ and τ.

TABLE 3

Bayes Estimates of the Waiting Time Distribution Parameters

(1.267, .466) 10.396 3.551	(.003, .435) 3.710 1.003	(3.135, .105) 2,631.507 22.977
(1.677, .418) 17.702 5.351	(1.965, .867) 12.697 7.133	(2.166, .131) 394.021 8.726
(.303, .442) 4.195 1.354	(1.642, .212) 54.849 5.168	(1.075, .598) 6.761 2.930

Comparison of Tables 1, 2, and 3 plus the prior estimates shows that (1) the posterior Bayes estimates effectively restrain the rather extreme fluctuation that afflict these small sample maximum likelihood estimates, and (2) the posterior Bayes estimates are overall slightly better than both the prior Bayes estimates and the maximum likelihood estimates.

Ignoring the observed terminal waiting time in state 2 of $w = 21.3$, the estimated transition probability matrix is

$$\underset{\sim}{P}^* = \begin{bmatrix} .318 & .455 & .227 \\ .180 & .520 & .300 \\ .082 & .306 & .612 \end{bmatrix}.$$

Incorporating the observed terminal waiting time according to the procedure of Section 7 requires that we calculate $F^{\tilde{\theta}}_{2j}(21.3)$ for $j=1,2,3$ to incorporate in the estimator given by equation (13). These calculated values are

$$F_{21} = .833, \quad F_{22} = .347, \quad \text{and } F_{23} = .639,$$

and hence the estimate $\tilde{\underset{\sim}{P}}$ is

$$\tilde{\underset{\sim}{P}} = \begin{bmatrix} .318 & .455 & .227 \\ .179 & .516 & .305 \\ .082 & .306 & .612 \end{bmatrix}.$$

This compares favorably with the maximum likelihood estimates and the prior estimates.

9. ACKNOWLEDGEMENTS

An earlier version of this report was presented at the National Meeting of the Operations Research Society of America, San Francisco, California, May 9, 1977. The author wishes to thank Mr. Che Fang for his assistance with this work and Professor Morris H. DeGroot for his comments.

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by ONR under Contract No. N00014-77-C-0095.

Any opinions, findings, conclusions or recommendations expressed herein are those of the author and do not necessarily reflect the views of the Defense Advanced Research Projects Agency, the Office of Naval Research, or Carnegie-Mellon University.

REFERENCES

[1] Aitchison, J. and J. A. C. Brown, The Lognormal Distribution, Cambridge: Cambridge University Press, 1969.

[2] Bartholomew, D. J., Stochastic Models for Social Processes, 2nd ed., New York: Wiley, 197[illegible].

[3] Brock, D. B., "Bayesian Analysis for Markov Renewal Processes," Metron, [illegible], [illegible], 1971.

[4] Box, G. E. P. and D. R. Cox, "An Analysis of Transformations," J. Roy. Statist. Soc., Series B, 26, 211, 1964.

[5] Cane, Violet R., "Behavior Sequences as Semi-Markov Chains," J. Roy. Statist. Soc., Series B, Vol. 21, 36-5[illegible], 1959.

[6] DeGroot, Morris H., Probability and Statistics, Reading, MA: Addison-Wesley, 1975.

[7] Ginsberg, Ralph B., "Semi-Markov Processes and Mobility," Journal of Mathematical Sociology, 1, pp. 233-262, July 1971.

[8] Ginsberg, Ralph B., "Critique of Probabilistic Models: Application of the Semi-Markov Model to Migration," Journal of Mathematical Sociology, 2, Vol. 1, pp. 63-82, 1972.

[9] Johnson, Norman L. and Samuel Kotz, Distributions in Statistics: Continuous Univariate Distributions-1, New York: Wiley, 1970.

[10] Johnson, Norman L. and Samuel Kotz, Distributions in Statistics: Continuous Multivariate Distributions, New York: Wiley, 1972.

[11] Levy, Paul, "Systémes Semi-Markoviens á au Plus une Infinité Dénombrable d'états Possibles," Proc. Int. Congr. Math., Amsterdam, Vol. 2, p. 294, 1954.

[12] Mann, N. R., Schafer, R. E., and Singpurwalla, N. D., Methods for Statistical Analysis of Reliability and Life Data, New York: John Wiley, 1974.

[13] Martin, J. J., Bayesian Decision Problems and Markov Chains, New York: Wiley, 1967.

[14] Moore, E. H. and R. Pyke, "Estimation of the Transition Distributions of a Markov Renewal Process," Annals of the Institute of Statistical Mathematics, 20, 411-424, 1968.

[15] Press, S. J., Applied Multivariate Analysis, New York: Holt, Rinehart, and Winston, 1972.

[16] Pyke, R., "Markov Renewal Processes with Finitely Many States," Annals of Mathematical Statistics, 32, pp. 1243-1259, 1961b.

[17] Pyke, R., "Markov Renewal Processes: Definitions and Preliminary Properties," Annals of Mathematical Statistics, 32, 1231-1242, 1961a.

[18] Smith, W. L., "Regenerative Stochastic Processes," Proc. Roy. Soc. (London), Ser. A., Vol. 232, 6-31, 1954 (abstract of this paper in Proc. Int. Congr. Math., Amsterdam, Vol. 2, pp. 304-305, 1955).

[19] Takács, Lajos, "Some Investigations Concerning Recurrent Stochastic Processes of a Certain Type," Magyar Tud. Akad. Mat. Kutato Int. Koz, Vol. 3, pp. 115-128, 1954.

[20] Wilson, E. B. and M. M. Hilferty, "The Distribution of Chi-Square," Proceedings of the National Academy of Sciences, Washington, 17, pp. 684-688, 1931.

DATE
ILME